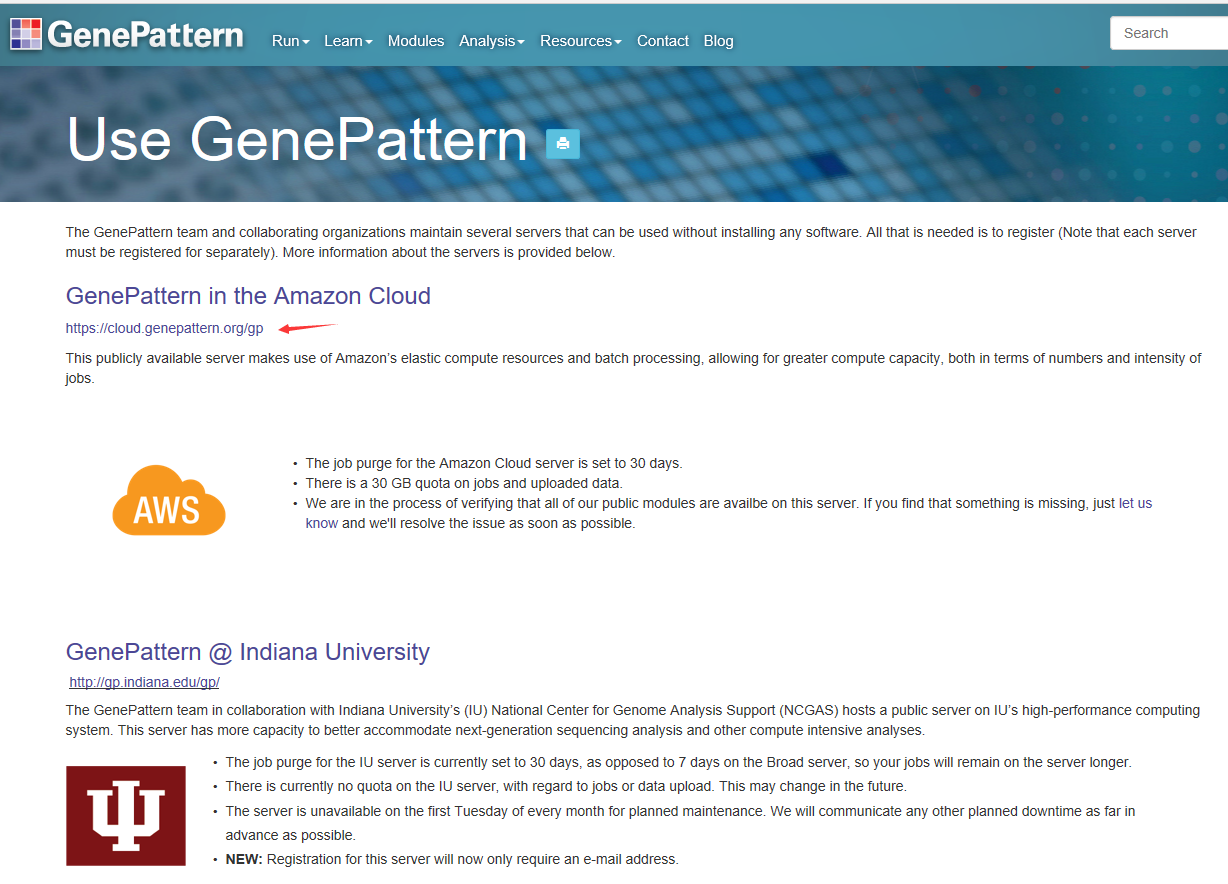
**submap使用教程**

**submap是一个用于比较表达谱相似性，从而做出推断的算法。比如训练集中A组和测试集B组在submap计算下显示出显著的相似性，那么A组可能和B有着类似的特性，这种特性可以体现在对治疗的响应性上。**

**目前没有计算submap的R软件包，其计算过程依赖于GenePattern云服务器中的submap模块，如下图所示：**



**网址为：**<https://cloud.genepattern.org/gp>

**或直接进入登录界面：**

<https://cloud.genepattern.org/gp/pages/login.jsf>

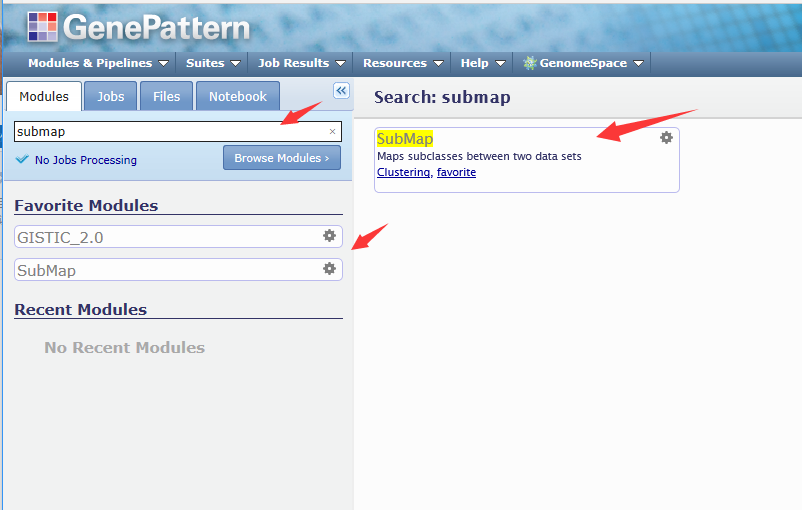
**使用前需要注册，如果在论文中涉及到submap算法的使用，请引用：**

**Hoshida Y, Brunet J-P, Tamayo P, Golub TR, and Mesirov JP (2007). Subclass mapping: identifying common subtypes in independent disease data sets. PLoS One 2, 1195–1203.**

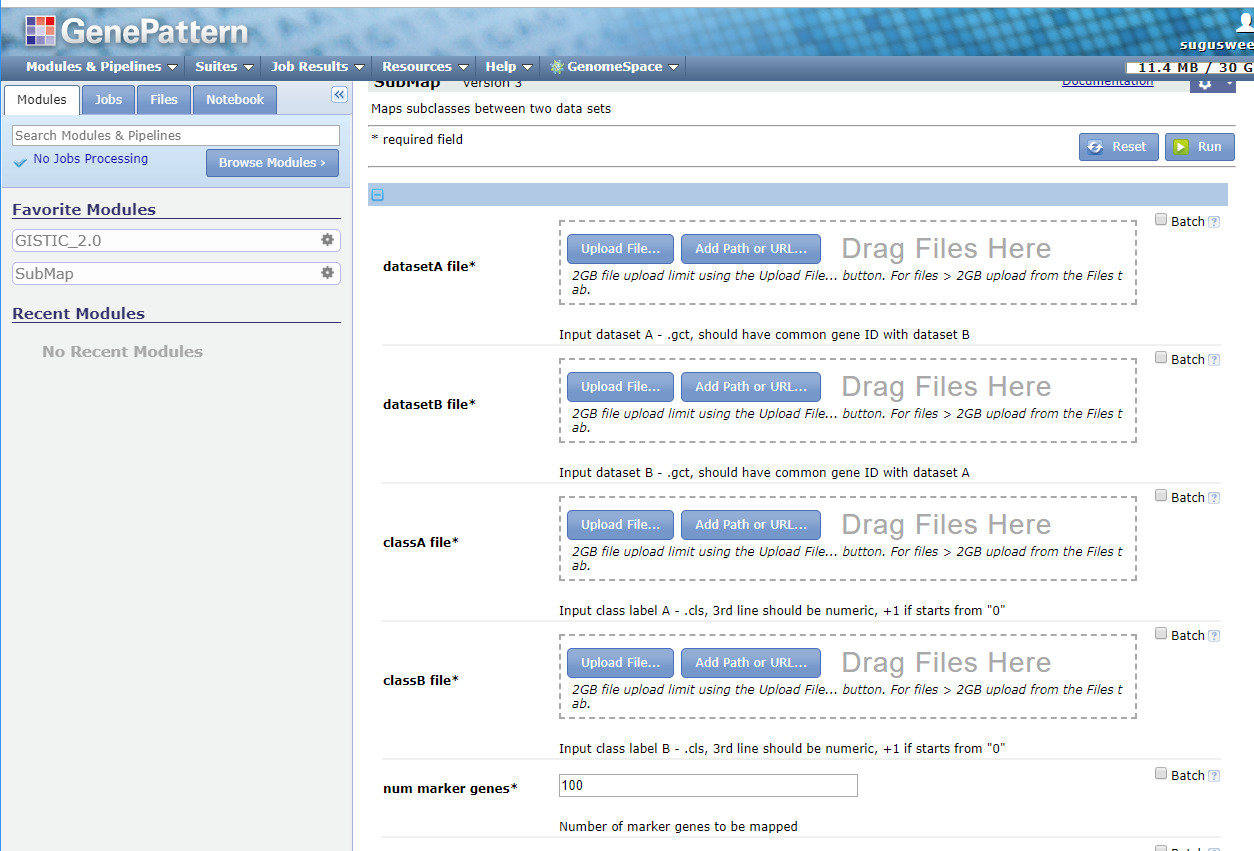
**另外，如果涉及使用本众筹包含的皮肤癌免疫治疗数据集，请引用：**

**Roh W, Chen P-L, Reuben A, Spencer CN, Prieto PA, Miller JP, Gopalakrishnan V, Wang F, Cooper ZA, Reddy SM (2017). Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. Science translational medicine 9: eaah3560**

**进入后如图所示，箭头所指是用户常用的模块，这里我常用submap和gistic，模块可以在Browse Modules里寻找，并拖动到左侧Favorite Modules，GenePattern可以实现非常多的生信分析过程，其难点是数据集的准备（格式），缺点是无法个性化调试一些细节，感兴趣的同学可以自行学习其他的模块：**



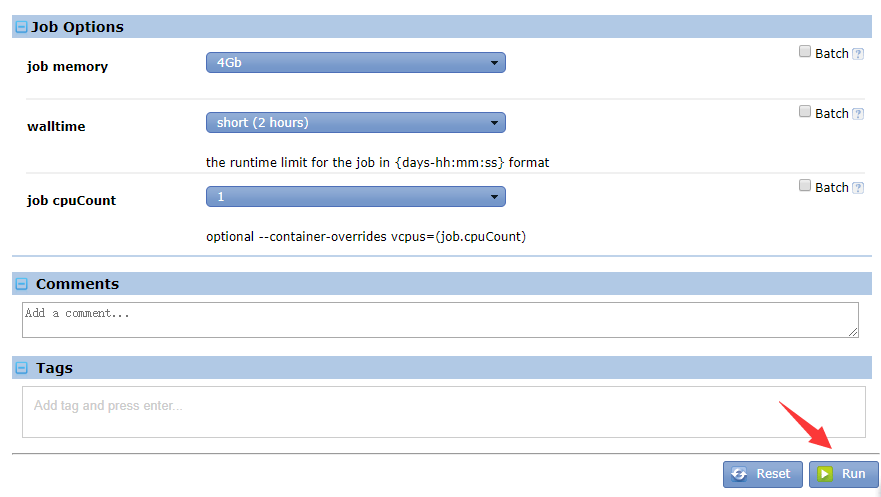
**点击左侧SubMap，进入数据输入界面，如下：**



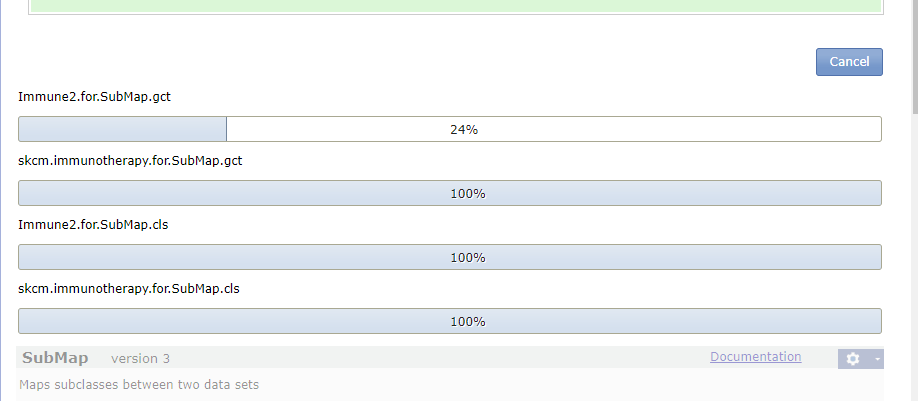
**输入数据选择如下，其他参数默认即可，P值校正可以采用Bonferroni和FDR，这里默认Bonferroni。**



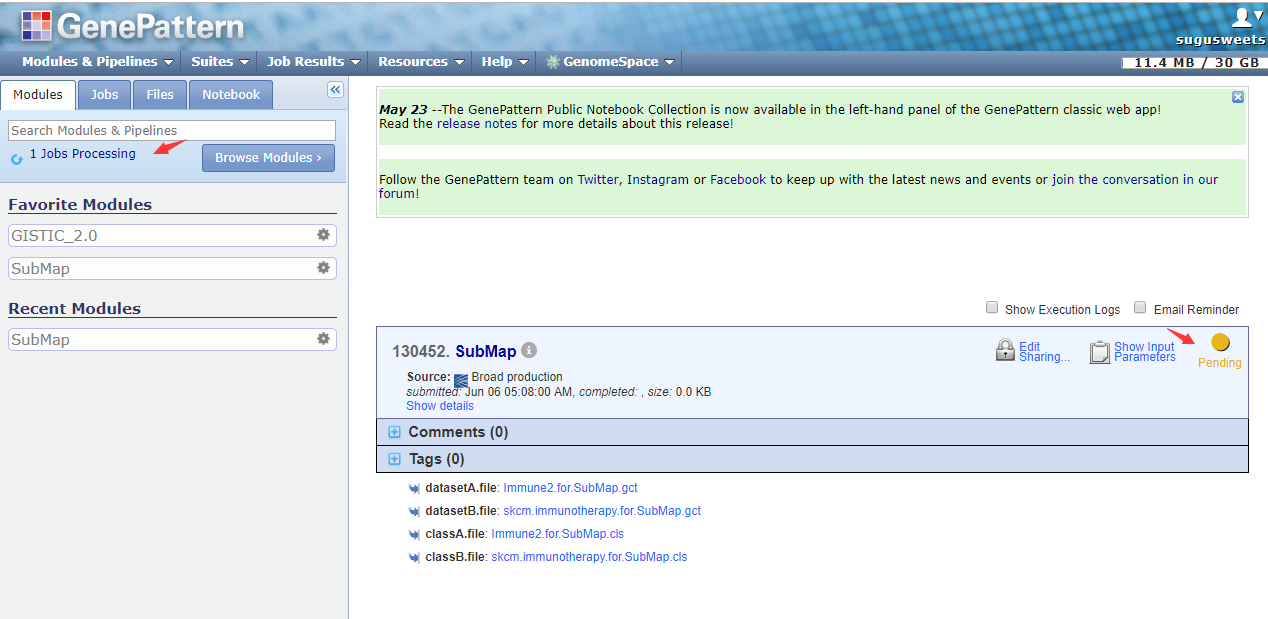
**然后翻到最下方可以选择使用的内存数，运行时间（最大1天，这也是在运行大样本量程序时该服务器的缺陷），随后提交运行，如下：**



**提交后请耐心等待数据上传，可滚动到页面上方查看进度，如下：**



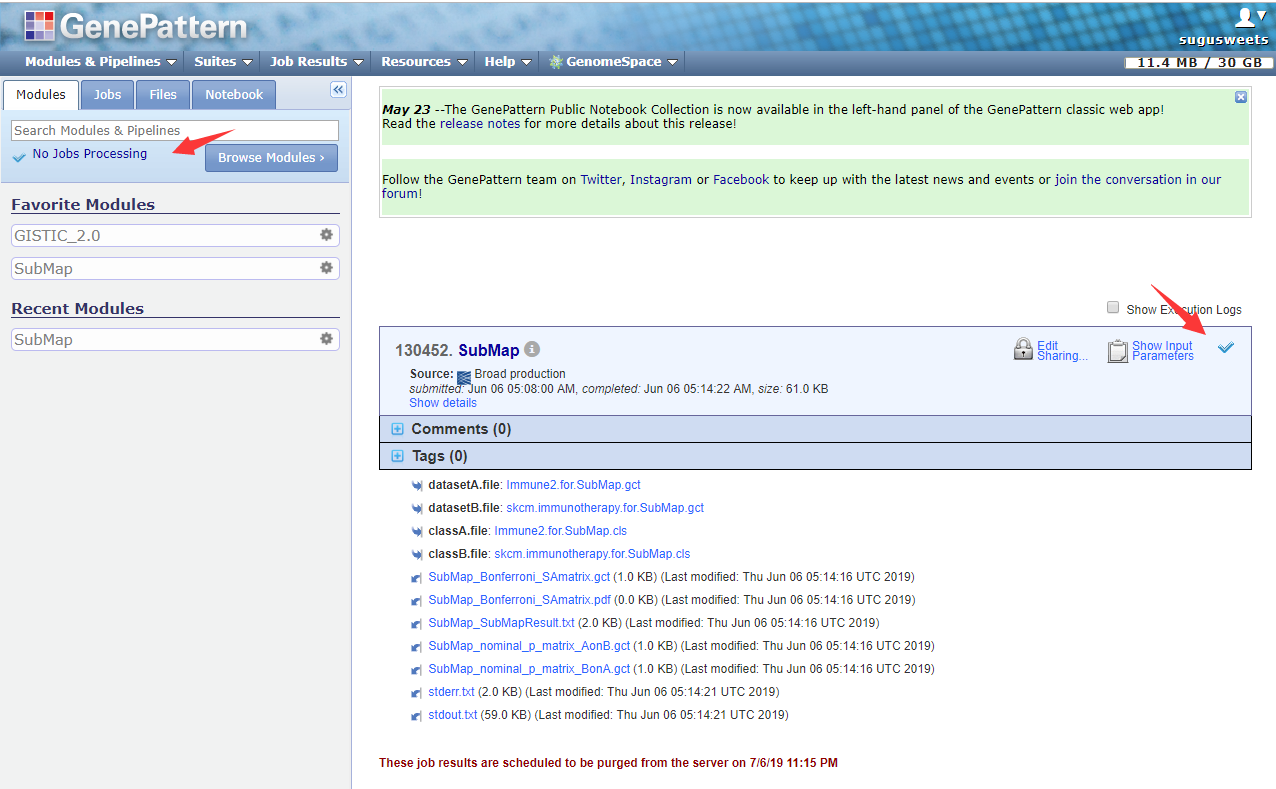
**上传成功后页面跳转，左侧提示有一个任务正在处理，右侧黄色pending表示进程在等待队列空位，如下：**



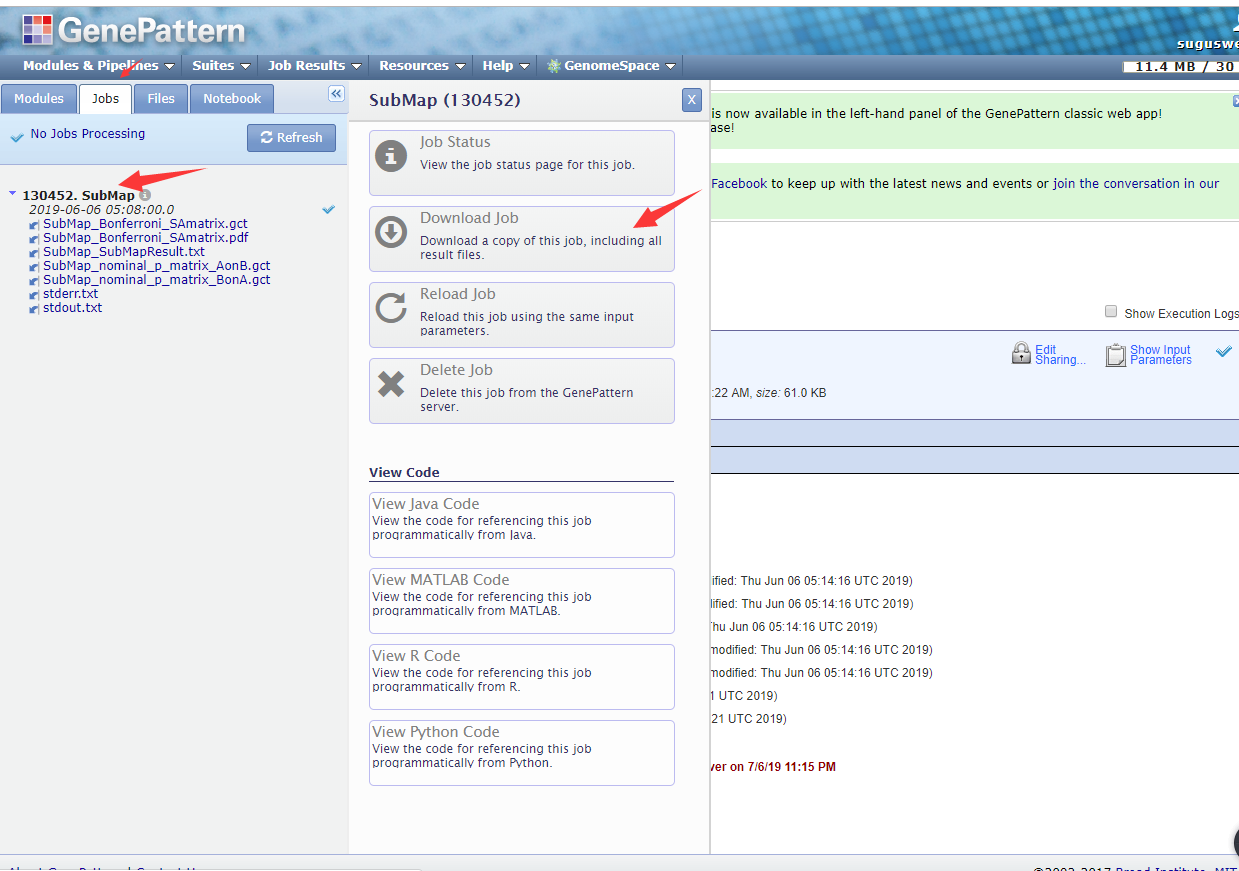
**提交程序后理论上可以关闭浏览器，开始运行后将变成蓝色的圈圈，请耐心等待；若失败则出现红色的叉叉，如下：**



**大概5分钟后，程序结束并生成结果，如下：**

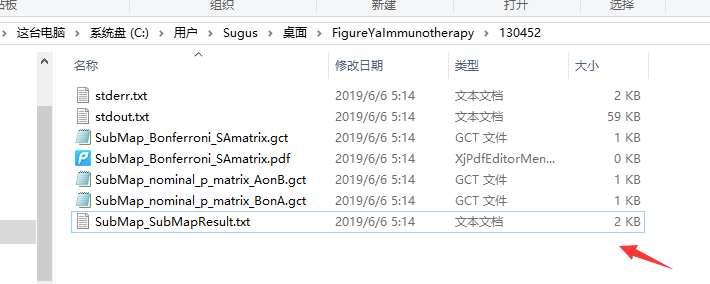


**随后依次点击Jobs，130452.subMap，Download Job即可下载结果，如下：**

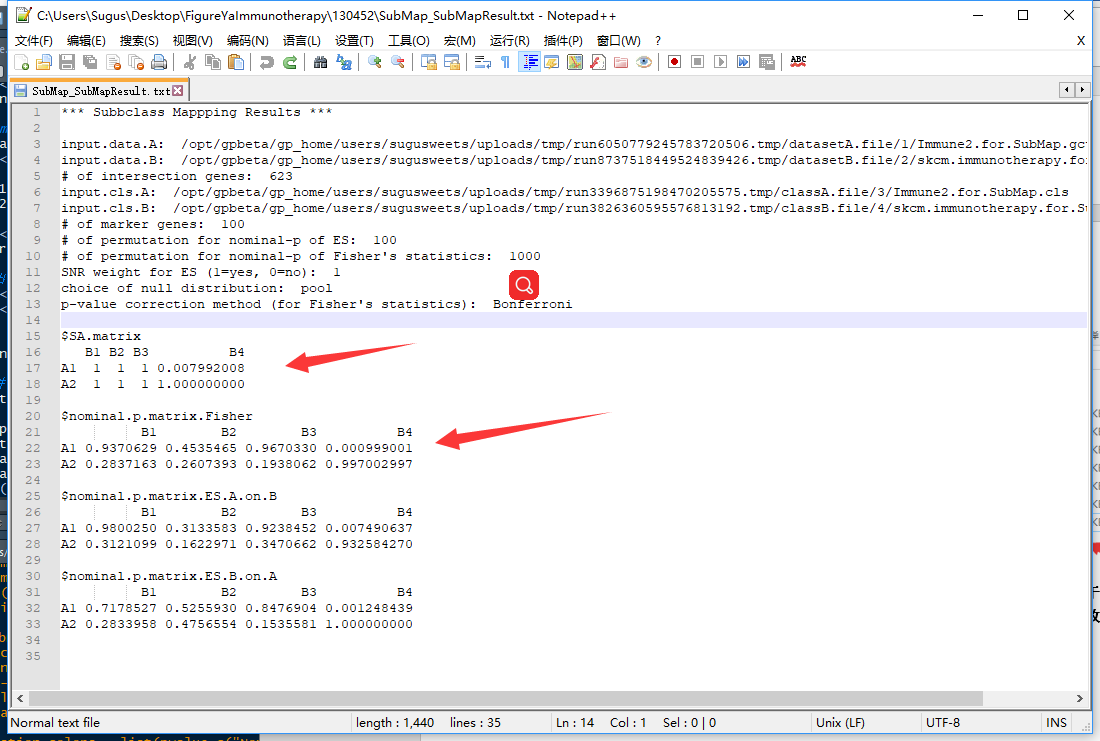


**注意130452是该Job的ID号，如果在使用此模块或其他模块产生了错误并且无法解决，请附带此ID至Google Group中GenePattern板块提问。**

**下载完成的压缩包130452.zip复制粘贴到文件夹中，并解压。其中SubMap\_Bonferroni\_SAmatrix.pdf其实就是p值的热图，但是自从GenePattern转到云服务器后，这个图就再也打不开了，所以只能手动绘制。用Notepad打开SubMap\_SubMapResult.txt，如下：**



**打开后，第一个箭头所指为Bonferroni校正的P值，第二个箭头为名义P值。由于Submap在运算中涉及种子和重抽样的过程，所以结果可能略有不同，关键结论理论上不会改变，如下：**



**记录这些值并回到R程序绘制热图，continue…**